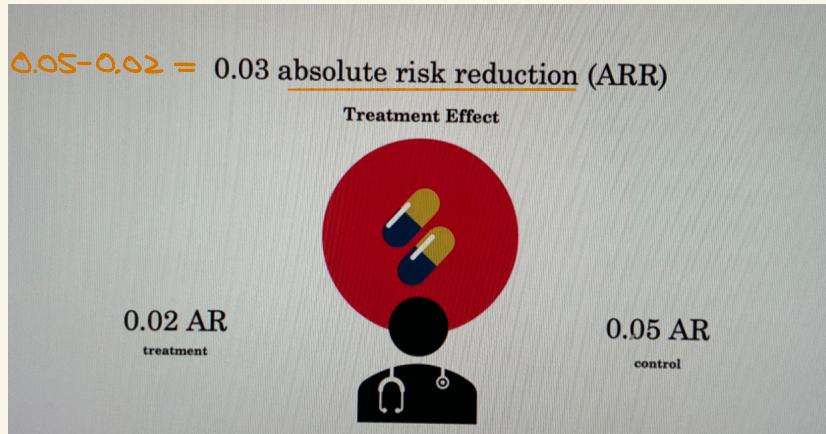
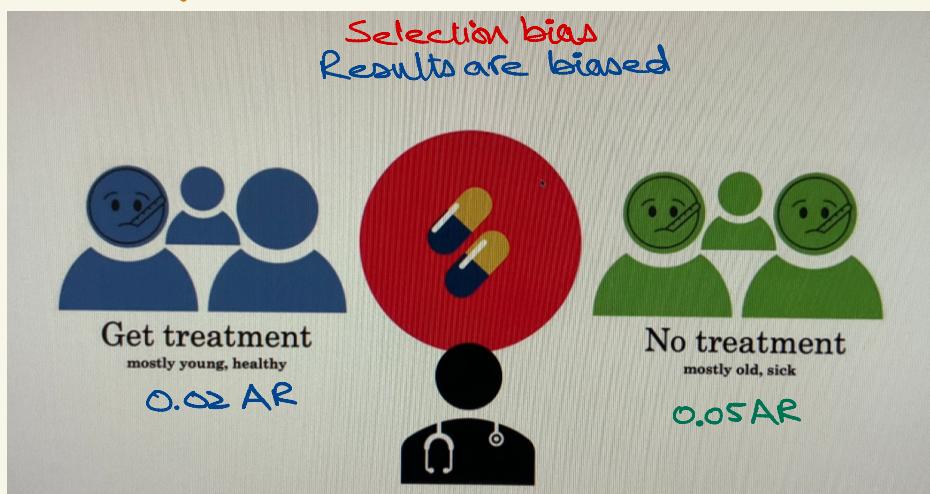


- With these two absolute risks, we can quantify the difference in risks.



Randomized Control Trials:

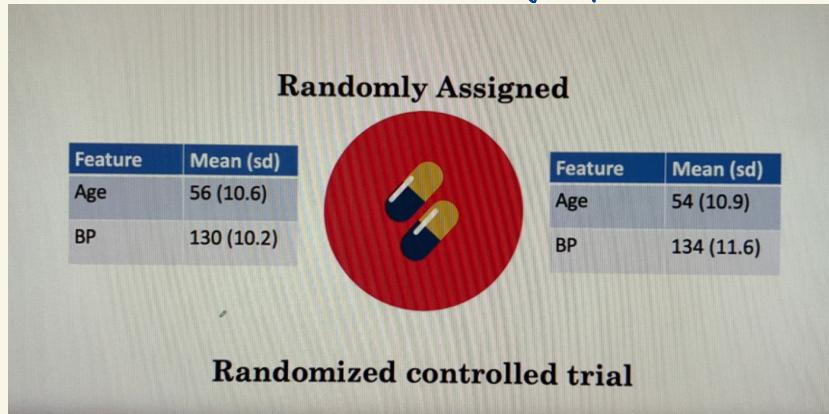
- Dr. is happy :) that we found an absolute risk reduction of 0.03. However, he is worried that; How do we select assignment to groups?



Absolute risk in treated group is smaller just because the young healthy patients are less likely to get a heart attack in the first place.

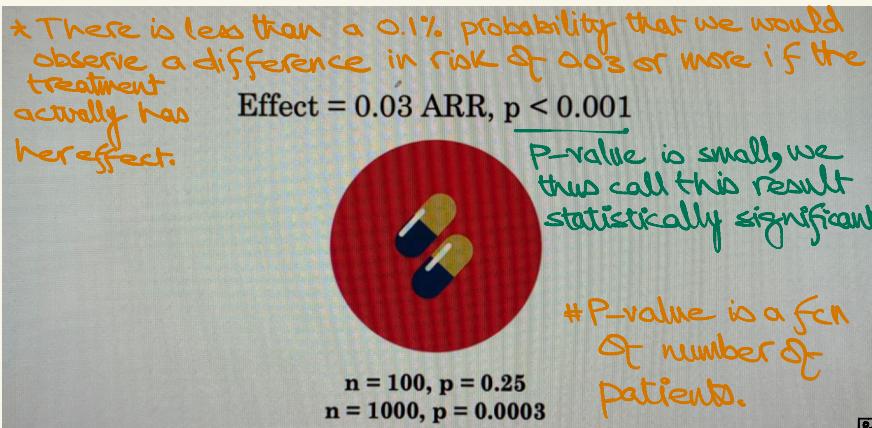
If our selection was biased then our treatment is not effective.

- The results will be effective, if we have randomly assigned the patients to treatment & control group.

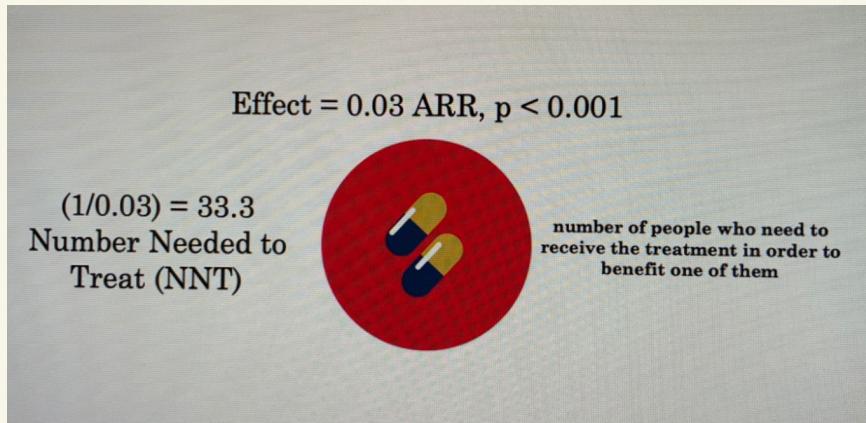


Randomized Control Trial (RCT): The setup of a medical experiment where we randomly allocate subjects to two or more groups, treat them differently and then compare them with respect to a measured response is called a randomized control trial, or RCT.

- The next question is to know about the significance of the effect; to confirm that the observed effect of the treatment, 0.03 ARR, was not by chance.
- To convey the statistical significance of the results, we report the P-value.



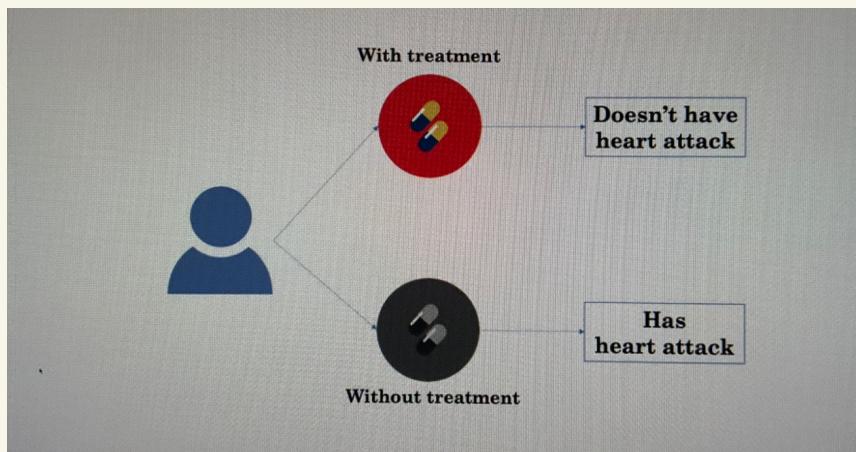
- The Absolute Risk Reduction (ARR) has a more interpretable def. in another related measure of treatment effect, which is the number needed to treat, or the NNT. The NNT is simply the reciprocal of the ARR.



In our case our NNT of 33.3 means that if we treat 33.3 people we will save one person from a heart attack.

Causal inference:

- Before giving a treatment to the patient the doctor wants to know; "What is the effect of a treatment on a patient?"



- We can show the potential outcomes in the table below:

| Unit | Outcome with treatment | Outcome without treatment | Effect |
|------|---------------------------|---------------------------|-----------|
| 1 | Doesn't have heart attack | Has heart attack | Benefit |
| 2 | Has heart attack | Has heart attack | No effect |
| 3 | Doesn't have heart attack | Doesn't have heart attack | No effect |
| 4 | Has heart attack | Doesn't have heart attack | Harm |

- We can represent these four possibilities for a patient's potential outcomes using the Neyman-Rubin causal model.

| i | Given | | $Y_i(1) - Y_i(0)$ |
|---|----------|----------|-------------------|
| | $Y_i(1)$ | $Y_i(0)$ | |
| 1 | 0 | 1 | -1 |
| 2 | 1 | 1 | 0 |
| 3 | 1 | 0 | 1 |
| 4 | 0 | 0 | 0 |

Doesn't have heart attack = 0
Has heart attack = 1

Neyman-Rubin causal model

- If we knew the potential outcome of many patients in a dataset, we can compute the mean for each column.

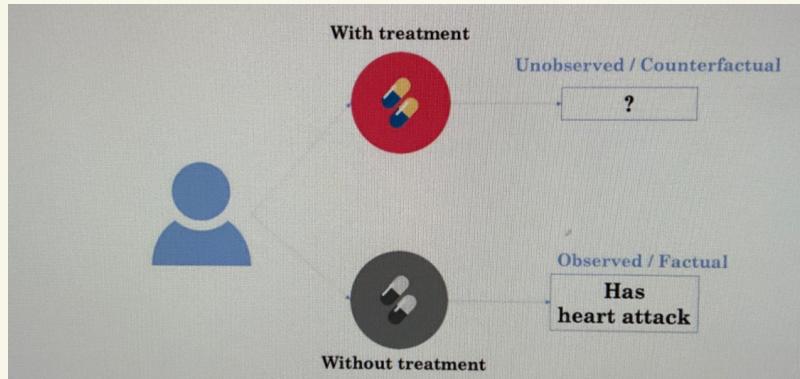
| $\mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$ | | | |
|---|----------|----------|-------------------|
| i | $Y_i(1)$ | $Y_i(0)$ | $Y_i(1) - Y_i(0)$ |
| 1 | 0 | 1 | -1 |
| 2 | 1 | 1 | 0 |
| 3 | 1 | 0 | 1 |
| 4 | 0 | 0 | 0 |
| 5 | 0 | 1 | -1 |
| Mean | 0.4 | 0.60 | <u>-0.20</u> |

Average Treatment Effect

- Average treatment effect is the expectation of the difference in the potential outcomes.

- But in reality we face the fundamental problem of causal inference. The challenge is we don't get to observe what happens to patient with and without treatment.

- Let's say for a patient in real life, we follow what happens to them when they don't get a treatment.



- Let's say we have a group of patients;

| | | Got Treated or not | $Y_i(1)$ | $Y_i(0)$ | $Y_i(1) - Y_i(0)$ |
|-----|-------|--------------------|-----------|----------|-------------------|
| i | W_i | | | | |
| 1 | 1 | | 0 | ? | ? |
| 2 | 0 | | ? Unknown | 1 | ? |
| 3 | 1 | | 1 | ? | ? |
| 4 | 1 | | 0 | ? | ? |
| 5 | 0 | | ? | 1 | ? |
| ... | ... | | ... | ... | ... |

How do we estimate Average Treatment Effect?
 $E[Y_i(1) - Y_i(0)]$

- It turns out that we can compute the average treatment effect when we are dealing with data from randomized control trials. The basic idea is to group the patients who get the treatment and the one who didn't. Then we take mean of their observed outcome.

- ATE (Average Treatment effect) is equal to -0.19.

- $ARR = -(\text{ATE})$
 $= 0.19$ (19% reduction)

| $E[Y_i(1) - Y_i(0)]$ | | |
|---------------------------------|-------|-------|
| In Randomized Controlled Trials | | |
| i | W_i | Y_i |
| 1 | 1 | 0 |
| 3 | 1 | 1 |
| 4 | 1 | 0 |
| ... | ... | ... |
| Mean | 0.32 | |

| i | W_i | Y_i |
|------|-------|-------|
| 2 | 0 | 1 |
| 5 | 0 | 1 |
| 6 | 0 | 0 |
| ... | ... | ... |
| Mean | 0.51 | |

$E[Y_i(1) - Y_i(0)] = E[Y_i|W = 1] - E[Y_i|W = 0]$
ATE earlier seen as ARR

- Q: Can we make a more individualized estimate?
Let's say we want to observe the effect of this treatment on a patient of age 56.

- This is called the conditional treatment effect.

| $\mathbb{E}[Y_i(1) - Y_i(0) \mid \text{Age} = 56]$ | | | |
|--|-----|-------|-------|
| In Randomized Controlled Trials | | | |
| $\mathbb{E}[Y_i(1) - Y_i(0) \mid \text{Age} = 56] = \mathbb{E}[Y_i \mid W = 1, \text{Age} = 56] - \mathbb{E}[Y_i \mid W = 0, \text{Age} = 56]$ | | | |
| i | Age | W_i | Y_i |
| 14 | 56 | 1 | 0 |
| 18 | 56 | 1 | 1 |
| Mean | | 0.5 | |
| i | Age | W_i | Y_i |
| 23 | 56 | 0 | 0 |
| Mean | | | 0 |

$\mathbb{E}[Y_i \mid W = 1, \text{Age} = 56] - \mathbb{E}[Y_i \mid W = 0, \text{Age} = 56]$

Very few examples to directly estimate from data

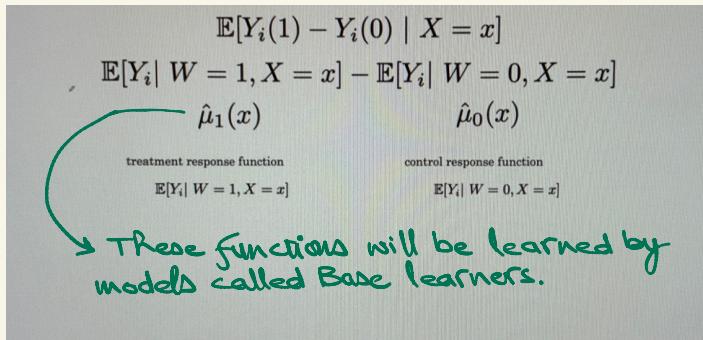
- Due to very few examples, we might not believe that we have a precise estimate. This can be even more of a problem, if we are not just considering age but also the BP etc. One solution to this problem could be that we learn the relationship between Age, BP and χ , then we could use the relationship to get the estimates.

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid \text{Age} = 56, \text{BP} = 130]$$

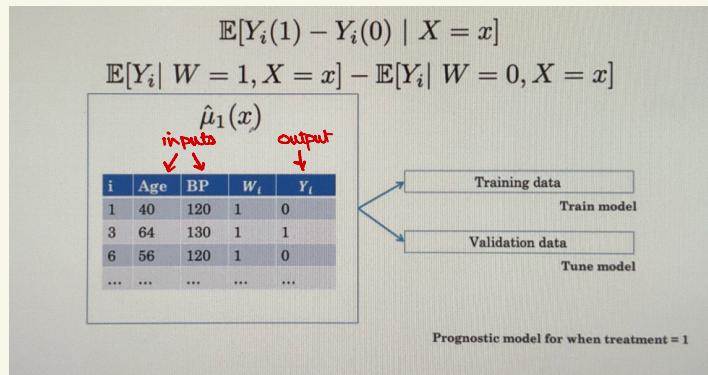
$$\mathbb{E}[Y_i(1) - Y_i(0) \mid X = x]$$

- * X represents the features like Age, BP, ...
- * x represents the values of these features.
 $x = [156, 60, \dots]$ is a vector

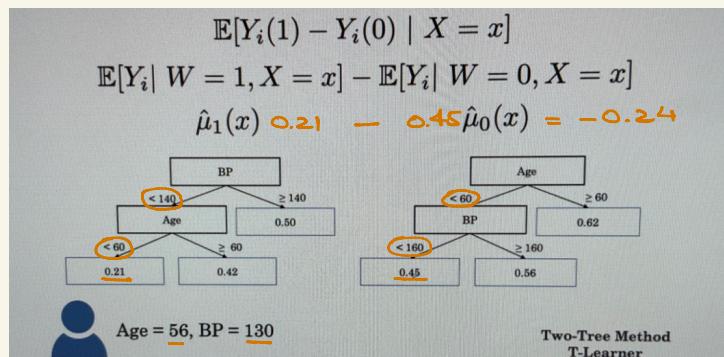
→ We found that we couldn't estimate it from the data. Instead we will use the function f_1 & f_0 to estimate this quantity.



- Base Learners can be for example Decision Tree or Linear models.

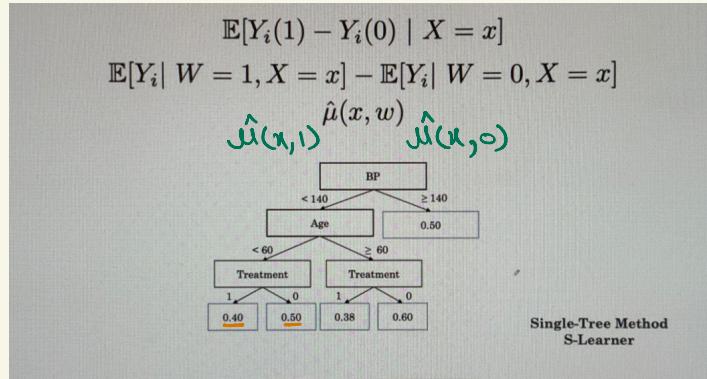


- These models are prognostic models; it gives the risk of an adverse event.
- Suppose we choose a Decision Tree model. We can estimate the risk for a new patient;



- The method of using these two models $\hat{\mu}_1$ & $\hat{\mu}_0$ and taking the difference to estimate the risk is called T-Learner.

- We can contrast T-Learner model with another method for making this estimation, and that is to use a single model $\hat{Y}(x, w)$. w can take the value of 0 or 1. This model is called S-Learner.



- Using this model we can estimate the treatment effect for a new patient. e.g. Age = 56, BP = 130.

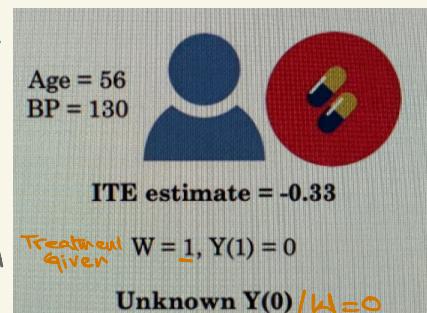
$$\begin{aligned}\hat{Y}(x, 1) - \hat{Y}(x, 0) &= 0.4 - 0.5 \\ &= -0.1 \text{ (Treatment Effect Estimate).}\end{aligned}$$

- Both of these learner methods have their disadvantages. For S-learner we might have tree that decide not to use the treatment feature. In T-learner as we have two models, each model is using half of the data. Hence, there is less data available to learn the relationships between the features.

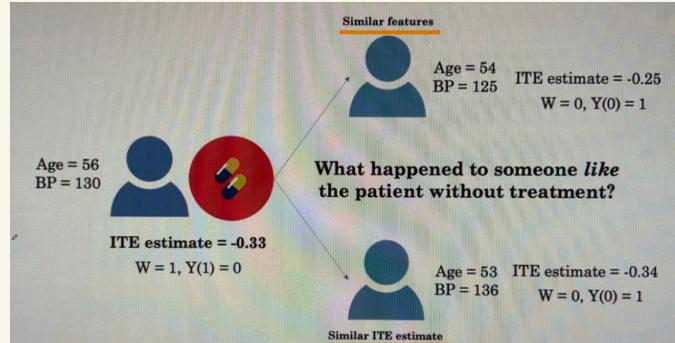
- Now the question is how to evaluate our estimator. How we will evaluate the ITE estimate?

So, for the patient in the figure on the right, if we want to evaluate the ITE we'd have to know the counterfactual. What would have happened to them without treatment.

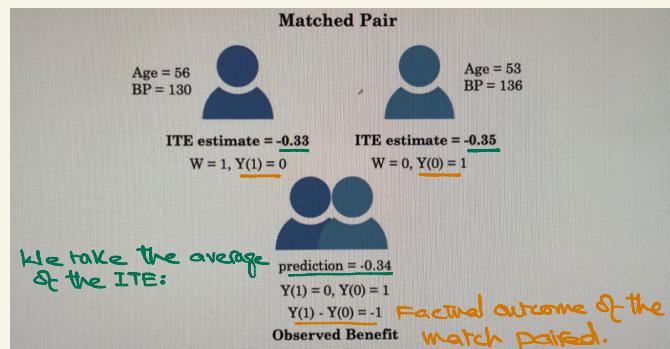
↳ To get $Y(0)$: Find out what happened to someone like the patient without treatment?



There are two ways which we can use to match this patient with another.



We can choose any one of the method. Here, we will choose the second one:

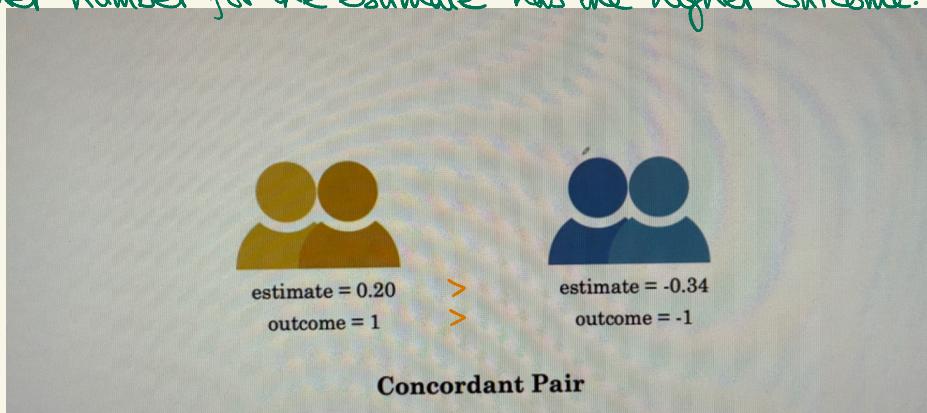


We can repeat this for other pairs of patients. Now we can determine; does higher predicted benefit actually correspond to observed benefit?

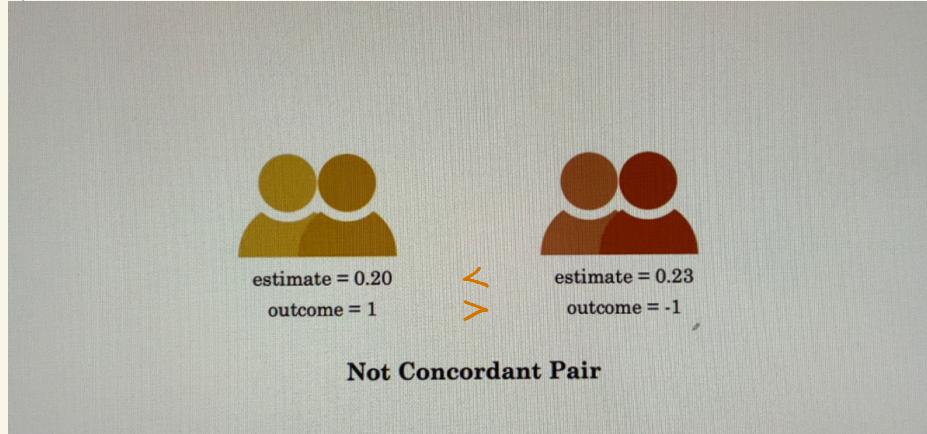
c-for-benefit method can be used to compute whether the higher predicted benefit actually corresponds to a higher observed benefit.



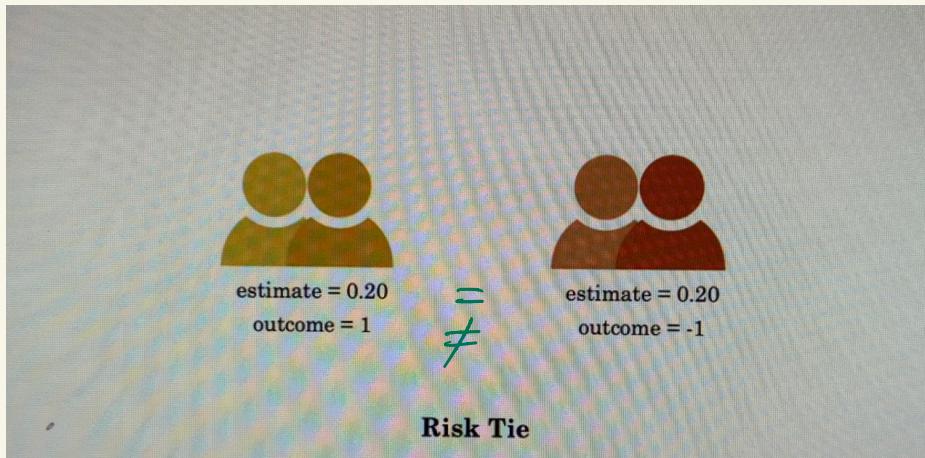
- This time instead of binary 0 and 1 outcomes or time to event outcomes. We have three possible outcomes, 1, -1 and 0.
- Some of the core concepts behind $\leq 2 \times$; This time we have pairs of pairs instead of pairs of individuals
- We call a pair of matched pairs concordant when the pair with the higher number for the estimate has the higher outcome.



- We call a pair of matched pairs not-concordant when the pair with the higher estimate number has the lower outcome level.

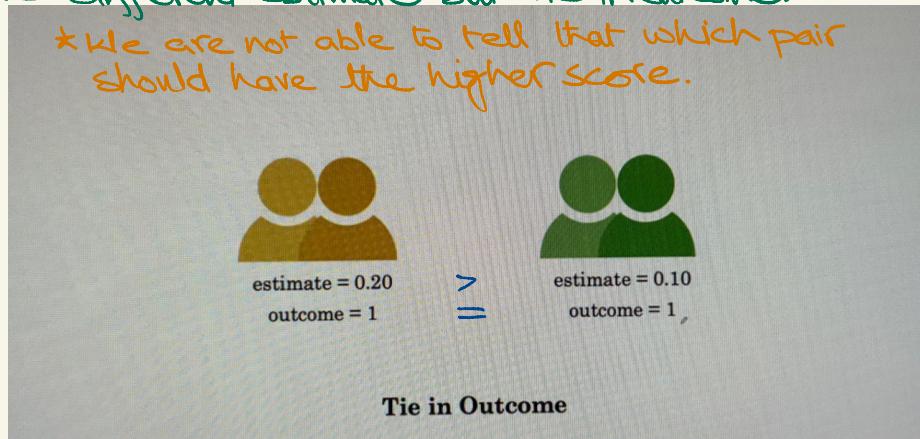


- We call a pair of matched pairs a risk tie when we get the same effect estimate for both pairs when they have different outcomes



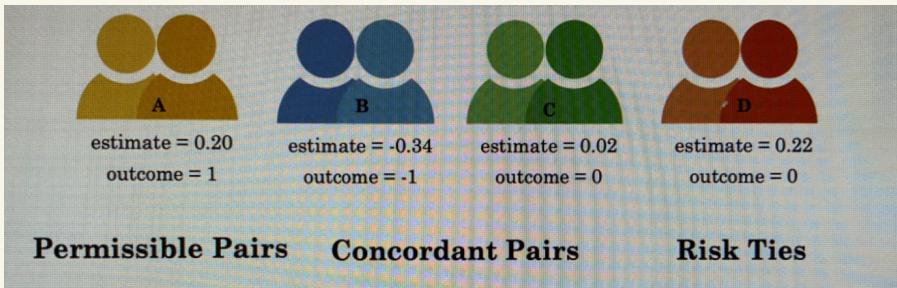
- We can't compare our effect estimates on pairs which have the different estimate but Tie in outcome.

* We are not able to tell that which pair should have the higher score.



Permissible pair: We can only compare pairs that have different observed outcomes.

$$C\text{-index} = \frac{\# \text{concordant pairs} + 0.5 \times \# \text{risk ties}}{\# \text{permissible pairs}}$$



- Let's compute the C-index with an example

Permissible Pairs Concordant Pairs Risk Ties

AB, AC, AD AB, AC, BC
 BC, BD

$$C\text{-index} = \frac{4 + 0.5 \times 0}{5}$$

$$= \frac{4}{5}$$

- Steps to process the data to perform the computation:

- 1) We split our RCT (Randomized Control Trials) data into two groups. (Got Treatment & Didn't get Treatment).
- 2) For each patient in both the groups, we use a learner to estimate the treatment effect (ITE)
- 3) Sort both groups by their treatment effect estimate, such that the patient we expect will have the most benefit will be at the top.
- 4) Then we can match patients by the rank such that the first patient in the treatment group gets matched with the first patient in

the control group.

| Match by Rank | | | | |
|---------------|-------|-------|-------|-------|
| i | X_i | W_i | Y_i | ITE |
| 34 | ... | 1 | 0 | -0.34 |
| 25 | ... | 1 | 1 | -0.32 |
| 63 | ... | 1 | 0 | -0.29 |
| ... | ... | ... | ... | ... |
| 32 | ... | 1 | 1 | 0.13 |

| Pair | YD | TE |
|----------|-----|--------|
| (34, 3) | -1 | -0.335 |
| (25, 54) | 0 | -0.325 |
| (63, 62) | -1 | -0.275 |
| ... | ... | ... |
| (32, 24) | 1 | 0.07 |

$$YD = Y_i^0 - Y_i^1$$

TE is the average of
ITE

5/ Now we can compute the C-Index using YD as outcome, and TE as the estimate

Q What does a C-for-Benefit 0.6 mean?

- The C-for-Benefit mean that give two randomly chosen pairs, A and B with different outcomes, what is the probability that the pair with the greater treatment effect estimate also has the greater Ydiff?

- 0.6 means the probability that the model correctly identified is 60%.

